On the inversion of an integral equation relating two wavefunctions in planes of an optical system suffering from an arbitrary number of aberrations

Hoenders, B.J.

Published in:
Optica Acta: International Journal of Optics

DOI:
10.1080/713820056

Publication date:
1979

OPTICA ACTA, 1979, VOL. 26, NO. 6, 711–730

On the inversion of an integral equation relating two wavefunctions in planes of an optical system suffering from an arbitrary number of aberrations

B. J. HOENDERS
Technical Physical Laboratories, State University at Groningen, Nijenborgh 18, 9747 AG Groningen, The Netherlands

(*Received 12 November 1977; revision received 23 February 1978*)

Abstract. If the wavefunction in the (not necessarily gaussian) image plane of an optical instrument is distorted by an arbitrary number of aberrations, the wavefunction in planes situated between the image plane and the plane of the specimen holder cannot be reconstructed by a Fourier series or a Fourier integral. This paper shows that the wavefunctions in those planes are uniquely determined by the values of the wavefunction in the image plane, and that an inversion formula, which in its structure is very similar to the well-known Fourier series, can be derived. Mathematically the problem concerns the inversion of the integral equation $h(x_1, x_2) = \int_\sigma \exp iS(x_1, x_2, y_1, y_2)\psi(y_1, y_1)\, dy_1 dy_2$ if the eikonal S is a multinomial.

1. Introduction

The rapid development of various techniques of image evaluation and image enhancement in both light and electron optics is due mainly to the applicability of the theory of Fourier analysis to problems concerning image formation and analysis, as first observed by Duffieux [1]. This is because if the process of image formation is restricted to the paraxial region the wavefunctions in the object plane, the plane of the exit pupil and the image plane are connected by spatial Fourier transforms in the case of *coherent illumination*, which will be assumed throughout this paper [2, §8.6.3].

Even more generally, Fourier analysis provides a powerful mathematical tool if, as is generally assumed, the image formation is *isoplanatic*, which means that the response (Green's) function of the optical system is invariant under translations perpendicular to the optical axis, since the various wavefunctions are then related by convolutions [2, §9.5.1].

A large number of fundamental problems can be treated by Fourier analysis, e.g. by using the Whittaker–Shannon sampling theorem. Felgett and Linfoot [3] calculated the information content of an image, and by using various ingenious methods Landé [4], in a very interesting paper, estimated the number of degrees of freedom of an image. (See also Wolter [5], Toraldo di Francia [6] and Hoenders [7] for alternative calculations and criticism concerning this concept.) Moreover, by using inverse Fourier transformation we can invert the imaging process, i.e. calculate the wavefunction in the exit pupil from the wavefunction in the image plane, etc., and hence solve the reconstruction problem, also known as the inverse diffraction problem. (For a free-space procedure see Lalor [8] and Shewell and Wolf [9].) The question then arises of how to modify all these results if, instead of making the paraxial or isoplanatic approximation, we consider the image formation of a coherently illuminated object imaged through any optical system and therefore allow

 0030-3909/79/2606 0711 $02·00

an arbitrary number of aberrations to blur the image. Because all the problems arising in Fourier optics are analysed by using the theory of Fourier series or the Fourier integral in some way, it seems very natural that, if we are to develop a theory of generalized Fourier optics to treat problems connected with blurred images, we must generalize the Fourier series for this case. Such a result has already been obtained by Ferwerda and Hoenders [10], who considered the case of one-dimensional imaging.

It is the aim of this paper to generalize the results obtained by Ferwerda and Hoenders and to obtain an inversion theorem for the general reconstruction (inverse diffraction) problem.

This problem can be stated as follows: is it possible to reverse the order of image formation in an optical instrument suffering from an *arbitrary* number of aberrations to calculate, for example, the wavefunction just behind the object from the values of the wavefunction in the image plane?

This reduces to the inversion of the integral equation of the first kind,

$$h(x_1, x_2) = \iint_\sigma \exp iS(x_1, x_2, y_1, y_2)\psi(y_1, y_2)\, dy_1 dy_2, \tag{1.1}$$

where $h(x_1, x_2)$ can be thought of as the wavefunction in the image plane, $\psi(y_1, y_2)$ as the wavefunction in a plane directly behind the object (object plane), the multinomial of order n $S(x_1, x_2, y_1, y_2)$ denotes the eikonal, and σ the aperture in the object plane.

It will be shown that if the eikonal S satisfies certain conditions, stated in lemma 2, ψ is uniquely determined by h and that the wavefunction ψ and h are related to each other by a generalized Fourier series:

$$\psi(\tau_1, \tau_2) = \lim_{c\to\infty} \sum_{j=1} \sum_{l=1} a_{l,j}(c) h(x_{1j}(c), x_{2l}(c)) \exp(-iS(x_{1j}, x_{2l}, \tau_1, \tau_2)), \tag{1.2}$$

where $a_{j,l}(c)$ are complex numbers, x_{1j} and x_{2l} are the roots of the transcendental equations $\exp(ix_j^n) + 1 = 0$, $j = 1, 2$, and the summation has to be extended over all numbers x_{1j} and x_{2l} such that $|x_{1j}| < c$ and $|x_{2l}| < \ln c$.

Equation (1.2) which, in its structure, is quite similar to the ordinary two-dimensional Fourier series (i.e. let $S = x_1 y_1 + x_2 y_2$), is (as is the Fourier series for bandlimited functions) of fundamental importance for an informational theoretical analysis of blurred images, as will be shown in a forthcoming paper.

2. Calculational procedure

We will first prove two lemmas needed for the proof of the main theorem. The first lemma gives some explicit results of the Riemann–Hilbert [11] problem, i.e. the construction of function which is analytic within a simply connected bounded domain and with prescribed values of its imaginary part on the boundary of the domain.

Lemma 1

Let the real function $I(\theta, c)$ (see figure 1) be defined on the interval $0 \leqslant \theta \leqslant 2\pi$ for all real values of $c > 0$ and, if n is a positive integer, let it be periodic with period $2\pi/n$ with respect to θ:

$$I(\theta, c) = \frac{\sin(nc^{-1/2})}{\exp(-c^{n+1})}\theta, \quad \text{if} \quad 0 \leqq \theta \leqq \exp(-c^{n+1}), \tag{2.1 a}$$

$$I(\theta, c)=\sin(nc^{-1/2}), \quad \text{if} \quad \exp(-c^{n+1}) \leqq \theta \leqq c^{-1/2}, \tag{2.1 b}$$

$$I(\theta, c)=\sin(n\theta), \quad \text{if} \quad c^{-1/2} \leqq \theta \leqq \frac{\pi}{n}-c^{-1/2}, \tag{2.1 c}$$

$$I(\theta, c)=\sin(nc^{-1/2}), \quad \text{if} \quad \frac{\pi}{n}-c^{-1/2} \leqq \theta \leqq \frac{\pi}{n}-\exp(-c^{n+1}), \tag{2.1 d}$$

$$I(\theta, c)=-\frac{\sin(nc^{-1/2})}{\exp(-c^{n+1})}\left(\theta-\frac{\pi}{n}\right), \quad \text{if} \quad \frac{\pi}{n}-\exp(-c^{n+1}) \leqq \theta \leqq \frac{\pi}{n} \tag{2.1 e}$$

and

$$I\left(\theta+l\frac{\pi}{n}, c\right)=(-1)^l g(\theta, c), \quad l=0, 1, 2, \ldots, 2n-1. \tag{2.1 f}$$

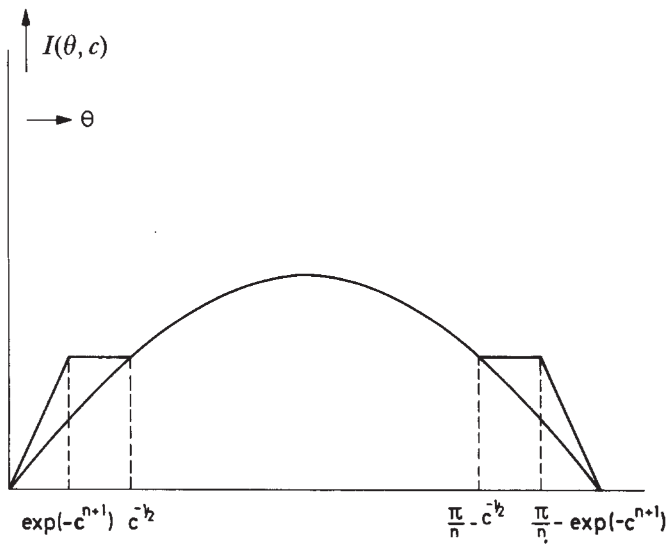

Figure 1. The function $I(\theta, c)$.

Then there exists a function $g(z, c)$, which is regular within the domain $|z|<1$ and continuous at and near the boundary $|z|=1$ with

$$\operatorname{Im}\{g(z, c); |z|=1\}=I(\theta, c), \tag{2.2}$$

and to every positive number ε there exists a positive number $C(\varepsilon)$ such that, for all numbers $c>C(\varepsilon)$,

$$|g(z, c)-z^n|<\varepsilon, \quad \text{if} \quad |z|\leqslant 1, \tag{2.3}$$

and $(g(z, c))^{1/n}$ is regular within the domain $|z|<1$.

Proof

The function $I(\theta, c)$ is periodic with period $2\pi/n$, and of bounded variation, and can therefore be expanded into a Fourier series:

$$I(\theta, c)=\sum_{j=0}^{\infty} a_j(c)\cos(jn\theta)+\sum_{j=0}^{\infty} b_j(c)\sin(jn\theta), \qquad -\pi\leqq\theta\leqq+\pi, \tag{2.4}$$

where

$$a_j(c)=\frac{1}{\pi}\int_{-\pi}^{+\pi} I(\theta, c)\cos(jn\theta)\,d\theta \tag{2.5}$$

and

$$b_j(c)=\frac{1}{\pi}\int_{-\pi}^{+\pi} I(\theta, c)\sin(jn\theta)\,d\theta. \tag{2.6}$$

Because $I(\theta, c)$ is odd with respect to $\theta=0$, we have that

$$a_j(c)\equiv 0, \quad \text{if} \quad j=0,1,2,\ldots. \tag{2.7}$$

We will also need to know the asymptotic behaviour of the numbers $b_j(c)$. Explicit evaluation of equations (2.5) and (2.6) or integration by parts leads to

$$b_j(c)=O\{c^{-1/2}j^{-2}\}, \quad j\neq 1, \tag{2.8}$$

where O denotes Landau's order symbol [12]. Since

$$\cos(jn\theta)=\operatorname{Im}\{ir^{jn}\exp(ijn\theta)\} \tag{2.9}$$

and

$$\sin(jn\theta)=\operatorname{Im}\{r^{jn}\exp(ijn\theta)\}, \quad \text{if} \quad r=1, \tag{2.10}$$

we have

$$I(\theta, c)=\operatorname{Im}\{g(z, c); |z|=1\}, \tag{2.11}$$

where

$$g(z, c)=\sum_{j=0}^{\infty} b_j(c)z^{jn} \tag{2.12}$$

and

$$z=r\exp(i\theta). \tag{2.13}$$

The power series on the right-hand side of equation (2.12) converges in the bounded and closed domain $|z|\leqq 1$ because its general term is $O\{c^{-1/2}j^{-2}\}$ and, by Abel's theorem, is continuous at and near the boundary $|z|=1$.

From equation (2.8) and Weierstrass's M test, we find that the power series (2.12) converges uniformly for all values of z within the *closed* domain $|z|\leqq 1$ and for all values of $c>0$, and that therefore summation and limit can be interchanged:

$$\lim_{c\to\infty} g(z, c)=\sum_{j=0}^{\infty}\{\lim_{c\to\infty} b_j(c)z^{jn}\}. \tag{2.14}$$

Observing that

$$\lim_{c\to\infty} b_1(c)=\frac{1}{\pi}\int_{-\pi}^{+\pi}\lim_{c\to\infty} I(\theta,c)\sin(n\theta)\,d\theta=1, \tag{2.15}$$

where the interchange between limit and integration is justified by Lebesque's dominated convergence theorem, and by using

$$\lim_{c\to\infty} b_j(c)=0, \quad j\neq 1, \tag{2.16}$$

combining equations (2.14), (2.15) and (2.16) leads to

$$\lim_{c\to\infty} g(z,c)=z^n, \tag{2.17}$$

uniformly for all values $|z|\leqq 1$ by Weierstrass's M test. Hence, by definition, for every positive number ε there exists a positive number $C(\varepsilon)$ such that, for all numbers $c>C(\varepsilon)$ and for all complex numbers $|z|\leqq 1$,

$$|g(z,c)-z^n|<\varepsilon. \tag{2.18}$$

However,

$$g(z,c)=z^n\{b_1(c)+\psi(z,c)\}, \tag{2.19}$$

where

$$\psi(z,c)=\sum_{j=2}^{\infty} b_j(c)z^{(j-1)n}, \tag{2.20}$$

and we therefore deduce from equation (2.18) that the function $\psi(z,c)$, which is analytic within the domain $|z|<1$, tends to zero if $c\to\infty$, uniformly for all values of $|z|\leqq 1$.

Therefore, because $b_1(c)$ tends to unity if $c\to\infty$, there exists a positive number C such that

$$|b_1(c)+\psi(z,c)|>0, \quad \text{if} \quad c>C \quad \text{and} \quad |z|\leqq 1, \tag{2.21}$$

i.e. the entire function $b_1(c)+\psi(z,c)$ has *no* zeros within the domain $|z|\leqq 1$ if $x>C$, which implies, recalling equation (2.19), that every branch of $(g(z,c))^{1/n}$ is an analytic function if $|z|<1$ and $c>C$.

We will now investigate the asymptotic behaviour of a certain double integral which appears frequently in the forthcoming analysis.

Lemma 2

Let

$$S(x_1,x_2,y_1,y_2)=\sum_{\substack{j,k,l,m\\ j+k+l+m\leqq n+1}}^{n} a_{j,k,l,m}\,{x_1}^j{x_2}^k{y_1}^l{y_2}^m, \tag{2.22}$$

where S is the eikonal, the numbers $a^{j,k,l,m}$ are arbitrary complex numbers such that $a_{n,0,1,0}$ and $a_{0,n,0,1}$ are *non-zero*, whereas $a_{n,0,0,1}$ and $a_{0,n,1,0}$ are zero [26], y_1 and y_2 are arbitrary real numbers, and x_1 and x_2 are arbitrary complex numbers with fixed moduli:

$$|x_1| = c \tag{2.23}$$

and

$$|x_2| = \ln c. \tag{2.24}$$

Then, if α, β, γ and δ are real numbers,

$$\int_\alpha^\beta dy_1 \int_\gamma^\delta dy_2 \exp iS(x_1, x_2, y_1, y_2)$$

$$= \gamma_1 (x_1 x_2)^{-n} \exp iS(x_1, x_2, y_1, y_2) \Bigg|_{y_1=\alpha}^{y_1=\beta} \Bigg|_{y_2=\gamma}^{y_2=\delta} \{1 + 0(1)\}, \tag{2.25 a}$$

if

$$(l-1)\pi + c^{-1/2} \leqq \arg (a_{n,0,1,0} x_1^n) \leqq l\pi - c^{-1/2},$$

$$(l-1)\pi + c^{-1/2} \leqq \arg (a_{0,n,0,1} x_2^n) \leqq l\pi - c^{-1/2}, \quad l = 1, 2, \ldots, 2n \tag{2.25 b}$$

and

$$\gamma_1 = (a_{n,0,1,0}\, a_{0,n,0,1})^{-1}. \tag{2.26}$$

The asymmetrical choice for the moduli of the complex numbers x_1 and x_2 and the construction of the function $g(z, c)$ will be explained immediately after the proof of this lemma.

The notation $\{1 + 0(1)\}$, which will be frequently used, is shorthand for 1 plus a function whose modulus tends to zero if c tends to infinity.

Proof

Using the identity

$$\exp iS(x_1, x_2, y_1, y_2) \equiv \gamma_1 \exp (iS(x_1, x_2, y_1, y_2) - ia_1 y_1 x_1^n - ia_2 y_2 x_2^n)$$

$$\times (x_1 x_2)^{-n} \frac{\partial^2}{\partial y_1 \partial y_2} \exp (ia_1 y_1 x_1^n + ia_2 y_2 x_2^n), \tag{2.27}$$

where

$$\left.\begin{aligned} a_1 &= a_{n,0,1,0} \\ \text{and} \qquad & \\ a_2 &= a_{0,n,0,1}, \end{aligned}\right\} \tag{2.28}$$

integration by parts leads to

$$\int_\alpha^\beta dy_1 \int_\gamma^\delta dy_2 \exp iS(x_1,x_2,y_1,y_2) = \gamma_1(x_1x_2)^{-n} \left[\exp iS(x_1,x_2,y_1,y_2)\Big|_{y_1=\alpha}^{y_1=\beta}\Big|_{y_2=\gamma}^{y_2=\delta} \right.$$

$$- \int_\alpha^\beta \exp(ia_1y_1x_1^n + ia_2y_2x_2^n)\frac{\partial}{\partial y_1}\{\exp(iS(x_1,x_2,y_1,y_2) - ia_1y_1x_1^n - ia_2y_2x_2^n)\}\,dy_1\Big|_{y_2=\gamma}^{y_2=\delta}$$

$$- \int_\gamma^\delta \exp(ia_1y_1x_1^n + ia_2y_2x_2^n)\frac{\partial}{\partial y_2}\{\exp(iS(x_1,x_2,y_1,y_2) - ia_1y_1x_1^n - ia_2y_2x_2^n)\}\,dy_2\Big|_{y_1=\alpha}^{y_1=\beta}$$

$$+ \int_\alpha^\beta dy_1 \int_\gamma^\delta dy_2 \exp(ia_1y_1x_1^n + ia_2y_2x_2^n)$$

$$\left. \times \frac{\partial^2}{\partial y_1 \partial y_2}\{\exp(iS(x_1,x_2,y_1,y_2) - ia_1y_1x_1^n - ia_2y_2x_2^n)\}\right]. \quad (2.29)$$

Then, integration by parts of the first and second integrals on the right-hand side of equation (2.29), using the identities

$$\exp(ia_1y_1x_1^n) = \frac{1}{ia_1x_1^n}\frac{\partial}{\partial y_1}\exp(ia_1y_1x_1^n) \quad (2.30)$$

and

$$\exp(ia_2y_1x_2^n) = \frac{1}{ia_2x_2^n}\frac{\partial}{\partial y_2}\exp(ia_2y_2x_2^n) \quad (2.31)$$

and applying the previously used transformation to the third integral on the right-hand side of equation (2.29), leads to the desired result.

Using the result of this lemma we can explain the asymmetrical choice $|x_1| = c$ and $|x_2| = \ln c$ and the application of the function $g(z, c)$. From the asymptotic expansion (2.25 *a*), we observe that the behaviour of the function on the left-hand side of equation (2.25 *a*) is determined *solely* by the term $a_1x_1^n$ of the eikonal S for values of arg $(a_1x_1^n)$ specified by equation (2.25), because for those values of $\arg(a_1x_1^n)$ the modulus of the function $\exp(ia_1x_1^n)$ is much larger or smaller than the modulus of the function $\exp(iS(x_1, x_2, y_1, y_2) - ia_1x_1^n)$.

This property is of paramount importance for the following analysis because the main trouble in developing a more dimensional theory, similar to the theory developed in [10], is the occurrence of mixed terms such as $x_1^{n-2}\,x_2^2$ in the eikonal which lead to a very complicated 'coupled' asymptotic behaviour of the eikonal. We will now explain the need for the function $g(z, c)$.

From the asymptotic expansion (2.25 *a*) we deduce the existence of so-called 'critical intervals' $s_l = \{(-c_1; l\pi - c^{-1/2} \leqq \arg(a_1x_1^n) \leqq l\pi + c^{-1/2},\ \ l = 1, 2, \ldots, 2n\}$ around the points $x_1 = \{x_1 = c, \arg(a_1x_1^n) = l\pi,\ l = 0, 1, \ldots, 2n\}$, where the asymptotic behaviour of the left-hand side of equation (2.25 *a*) is no longer determined by the term $a_1x_1^n$ of the eikonal, but in a very complicated way by the various other terms of the eikonal; i.e. there exist intervals on the circle $|x_1| = c$ where the asymptotic behaviour of equation (2.25 *a*) is not determined *solely* by $a_1x_1^n$.

However, as we will use only the left-hand side of equation (2.25 a) while integrating over x_1 and x_2, the contribution of these intervals to the integrals could be made negligible if we were able in some way to decrease the length of the critical intervals. This is achieved by introducing the function $g(x_1, c)$. Let us make the substitution $c^{-1}x_1 = (g(x_1, c))^{1/n}$, $|x_1| = c$, $0 \leqq \arg x_1 \leqq 2\pi$ in equation (2.25 a).

We then observe, recalling the definitions (2.1 a) and (2.1 b), that $\mathrm{Im}\,(a_1 c^n g(x_1, c))$ is much larger than the imaginary parts of the other terms of the eikonal everywhere on the circle $|x_1| = 1$, with the exclusion of intervals of width $\exp(-c^{n+1})$ around the critical points $x_l = \exp(i\phi_l)$, $\phi_l = 0, \pi/n, 2\pi/n, \ldots, 2\pi - \pi/n$. Hence, by introducing the function $g(x_1, c)$ we decreased the length of the critical intervals in such a way that their contribution by the integration of x_1 will vanish if $c \to \infty$ and, in a way, we will have to consider only those intervals where the asymptotic behaviour of equation (2.25 a) is determined by $\mathrm{Im}\,(a_1 c^n g(x_1, c))$.

We are now prepared to formulate the basic theorem of our theory, which is a generalization of the well-known formula, basic to Fourier optics,

$$\frac{2}{\pi} \int_0^{\tau} \sum_{n=-\infty}^{+\infty} \exp(in(x-y))\, dy = U(\tau - x). \tag{2.32}$$

Theorem 1

Let ξ, η, τ_1 and τ_2 denote real numbers with values within the interval [0, 1]. Then, if the numbers x_{1j} and $x_{2j'}$ are the roots of the transcendental equations

$$\omega(c, x_{1j}) \equiv \exp(ic^n x_{1j}^n) + 1 = 0 \tag{3.1}$$

and

$$\omega(\ln c, x_{2j'}) = \exp(i \ln^n c x_{2j'}^n) + 1 = 0, \tag{3.2}$$

$$\begin{aligned} \lim_{c \to \infty} \sum_j \sum_{j'} & (g(x_{1j}, c) g(x_{2j'}, c))^{1-1/n} (g'(x_{1j}, c) g'(x_{2j'}, c))^{-1} \\ & \times \int_0^{\xi} dy_1 \int_0^{\eta} dy_2 \exp\{iS(cx_{1j}, \ln(c) x_{2j'}, y_1, y_2) \\ & - iS(cx_{1j}, \ln(c) x_{2j'}, \tau_1, \tau_2)\} \\ & = U(\xi - \tau_1) U(\eta - \tau_2), \end{aligned} \tag{3.3}$$

where the summations are extended over all values of x_{1j} and $x_{2j'}$, with modulus not exceeding unity, $U(x)$ denotes Heaviside's unit step function, $g'(x, c) = \frac{\partial}{\partial x} g(x, c)$ and the Tannery series on the left-hand side of equation (3.3) converges uniformly for all values of ξ, η, τ_1 and τ_2 within the interval [0, 1]. (A Tannery series is a series whose general term depends explicitly on the number of terms of the series, for example $\sum_{j=1}^{n} \ln(1 - 1/jn)$ and $\sum_{j=1}^{n} (-1) j (1/j^2 + n)$ are Tannery series. The last series converges if n is an even number, although its general term tends to infinity!)

Proof

As typical examples we will consider the cases $\tau_1 < \xi$, $\tau_2 > \eta$ and $\tau_1 < \xi$, $\tau_2 < \eta$. All other possible locations of the point (τ_1, τ_2) can be treated completely analogously

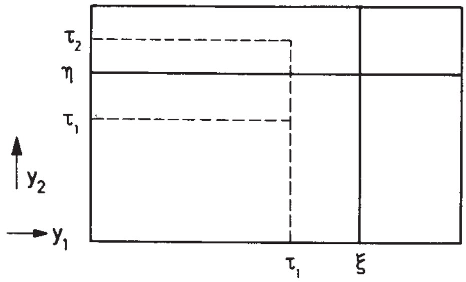

Figure 2. Positions of the points ξ, η, τ_1 and τ_2.

(see figure 2). In lemma 1 it was proved that the set of entire functions $g^{1/n}(z, c)$, entire for all values of z such that $|z|<1$ and defined at the boundary $|z|=1$, converge *uniformly* to z if c tends to infinity. The domain $D\{D: z; |z|<1\}$ is therefore mapped conformly onto the domain D' $\{D': w=g^{1/n}(z, c), |z|<1\}$ and the boundary $|z|=1$ of the domain D is mapped uniquely onto the boundary of the domain D' for sufficiently large values of c.

This knowledge will enable us to construct contours c_1 and c_2 in the complex $w_1=g^{1/n}(x_1, c)$ and $w_2=g^{1/n}(x_2, c)$ planes respectively such that they pass between two successive zeros of the functions $\omega(c, x_1)$ and $\omega(\ln(c), x_2)$. The reason for doing this is that the reciprocal of these functions are then uniformly bounded on the

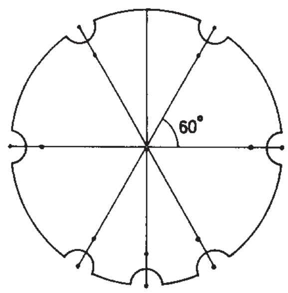

Figure 3. The contour c_1' for $n=3$.

contours c_1 and c_2 for all values of c. The contour c_1 is constructed as follows. Draw the curve $|x_1|=1$ and draw around each zero x_n of the function $\omega(c, x_1)$ a circle whose radius is half the distance to the next-nearest zero. Then, if the contour c_1' consists of parts of the circle $|x_2|=1$ and those parts of the circles around the points x_n which are intersected by the curve $|x_1|=1$, the contour c_1 is defined as the mapping of c_1' from the x_1 and w_1 planes. The contour c_2 is defined in a similar way.

Consider the contour integral

$$I(\xi, \eta, \tau_1, \tau_2, c) = -\frac{1}{4\pi^2} \oint_{x_1 \in c_1} dx_1 \oint_{x_2 \in c_2} dx_2 \, (c \ln(c))^n (g(x_1, c) g(x_2, c))^{1-1/n}$$
$$\times (\omega(c, g(x_1, c)) \omega(\ln c, g(x_2, c)))^{-1} \int_0^{\xi} dy_1 \int_0^{\eta} dy_2 \, K(x_1, x_2, y_1, y_2, \tau_1, \tau_2, c), \quad (3.4)$$

where

$$K(x_1, x_2, y_1, y_2, \tau_1, \tau_2, c) = \exp\{iS(cg^{1/n}(x_1, c), \ln(c) g^{1/n}(x_2, c), y_1, y_2)$$
$$- iS(cg^{1/n}(x_1, c), \ln(c) g^{1/n}(x_2, c), \tau_1, \tau_2)\}. \quad (3.5)$$

From the theorem of residues we find that equation (3.4) is equal to the series (3.3). The theorem of residues can be used [13] because, according to lemma 1, there exists a positive number C such that for all values $c > C$ the function $g^{1/n}(z, c)$ is entire within the domain $|z| < 1$ and continuous at and near the boundary $|z| = 1$, which implies that the integrand of the integral (3.4) satisfies the conditions of the theorem of residues formulated by Watson [13].

Furthermore, in evaluating the integral (3.4) we assumed that the roots of equations (3.1) and (3.2) are simple. The truth of this conjecture follows immediately from the observation that the roots of the equation $\exp(ic^n z^n) + 1 = 0$ are simple and that, according to lemma 2, from a certain number c the function $g^{1/n}(z, c)$ maps the domain $|z| < 1$ conformly and therefore uniquely onto the domain $w = g^{1/n}(z, c)$ because the function $g^{1/n}(z, c)$ converges *uniformly* to z for all numbers z with modulus $|z| \leqq 1$, if c tends to infinity.

As in the previous paper [10], we will evaluate the integral (3.4) using the asymptotic behaviour of the integrand on the contour for large values of c. We therefore introduce the functions

$$h_1(x_1, x_2, \tau_1, \tau_2, \eta, c) = \int_0^{\tau_1} dy_1 \int_0^{\eta} dy_2\, K(x_1, x_2, y_1, y_2, \tau_1, \tau_2, c), \tag{3.6}$$

and

$$h_2(x_1, x_2, \tau_1, \tau_2, \xi, \eta, c) = \int_{\tau_1}^{\xi} dy_1 \int_0^{\eta} dy_2\, K(x_1, x_2, y_1, y_2, \tau_1, \tau_2, c). \tag{3.7}$$

The asymptotic behaviour for $c \to \infty$ of h_1 and h_2 at the contours $|x_1| = 1$ and $|x_2| = 1$, with the exclusion of small intervals of width $\exp(-c^{n+1})$ symmetric around the points $x_1 = \exp(il\pi/n)$ and $x_2 = \exp(il\pi/n)$, $l = 0, 1, 2, \ldots, 2n-1$ and $l' = 0, 1, 2, \ldots, 2n-1$, is obtained from lemma 2 and the construction (see equations (2.1 *a*), (2.1 *b*) and (2.2)) of the function $g(z, c)$:

$$h_1(x_1, x_2, \tau_1, \tau_2, \eta, c) = \gamma_1(c \ln(c))^{-n}(g(x_1, c)g(x_2, c))^{-1}$$
$$\times K(x_1, x_2, y_1, y_2, \tau_1, \tau_2, c) \Bigg|_{y_1=0}^{y_1=\tau_1} \Bigg|_{y_2=0}^{y_2=\eta} \{1 + 0(1)\}$$

if

$$l\frac{\pi}{n} + \exp(-c^{n+1}) \leqq \arg x_1 \leqq (l+1)\frac{\pi}{n} - \exp(-c^{n+1}),$$
$$l\frac{\pi}{n} + \exp(-c^{n+1}) \leqq \arg x_2 \leqq (l+1)\frac{\pi}{n} - \exp(-c^{n+1}), \quad l = 0, 1, 2, \ldots, 2n-1 \tag{3.8}$$

and

$$h_2(x_1, x_2, \tau_1, \tau_2, \xi, \eta, c) = \gamma_1(c \ln c)^{-n}(g(x_1, c)g(x_2, c))^{-1}$$
$$\times K(x_1, x_2, y_1, y_2, \tau_1, \tau_2, c) \Bigg|_{y_1=\tau_1}^{y_1=\xi} \Bigg|_{y_2=0}^{y_2=\eta} \{1 + 0(1)\}$$

if

$$l\frac{\pi}{n} + \exp(-c^{n+1}) \leqq \arg x_1 \leqq (l+1)\frac{\pi}{n} - \exp(-c^{n+1}),$$
$$l\frac{\pi}{n} + \exp(-c^{n+1}) \leqq \arg x_2 \leqq (l+1)\frac{\pi}{n} - \exp(-c^{n+1}), \quad l = 0, 1, 2, \ldots, 2n-1 \tag{3.9}$$

From equation (2.11) and the definitions (3.1) and (3.2), we obtain the asymptotic behaviour of the functions $\omega(c, x_1)$ and $\omega(\ln c, x_2)$. Recalling the construction of the contours c_1 and c_2 (see figure 3), which was performed in such a way that the functions $(\omega(c, g(x_1, c)))^{-1}$ and $(\omega(\ln c, g(x_2, c)))^{-1}$ are *uniformly* bounded on the contours for all values of c, and by using equation (2.11) together with the definitions (3.1) and (3.2), we derive the asymptotic behaviour of the functions $(\omega(c, g(x_1, c)))^{-1}$ and $(\omega(\ln c, g(x_2, c)))^{-1}$ as

$$(\omega(a_j, g(x_j, c)))^{-1} = 1 + 0\{\exp(-a_j^n \sin(n \arg x_j))\},$$

if

$$2l\frac{\pi}{n} + \frac{c^{-1/2}}{n} \leqq \arg x_j \leqq (2l+1)\frac{\pi}{n} - \frac{c^{-1/2}}{n}, \quad \text{and} \quad x_j \in c_j, \tag{3.10 a}$$

$$= 1 + 0\{\exp(-n a_j^n c^{-1/2})\},$$

if

$$(2l+1)\frac{\pi}{n} - \frac{c^{-1/2}}{n} \leqq \arg x_j \leqq (2l+1)\frac{\pi}{n} - \frac{\exp(-c^{n+1})}{n},$$

$$2l\frac{\pi}{n} + \frac{\exp(-c^{n+1})}{n} \leqq \arg x_j \leqq 2l\frac{\pi}{n} + \frac{c^{-1/2}}{n}, \quad \text{and} \quad x_j \in c_j, \tag{3.10 b}$$

$$= 0 + 0\{\exp(a_j^n \sin(n \arg x_j))\},$$

if

$$(2l+1)\frac{\pi}{n} + \frac{c^{-1/2}}{n} \leqq \arg x_j \leqq (2l+2)\frac{\pi}{n} - \frac{c^{-1/2}}{n}, \quad \text{and} \quad x_j \in c_j, \tag{3.10 c}$$

$$= 0 + 0\{\exp(-n a_j^n c^{-1/2})\},$$

if

$$(2l+1)\frac{\pi}{n} + \frac{\exp(-c^{n+1})}{n} \leqq \arg x_j \leqq (2l+1)\frac{\pi}{n} + \frac{c^{-1/2}}{n},$$

$$(2l+2)\frac{\pi}{n} - \frac{c^{-1/2}}{n} \leqq \arg x_j \leqq (2l+2)\frac{\pi}{n} - \frac{\exp(-c^{n+1})}{n} \quad \text{and} \quad x_j \in c_j, \tag{3.10 d}$$

$$= 0(1),$$

if

$$l\frac{\pi}{n} - \frac{\exp(-c^{n+1})}{n} \leqq \arg x_j \leqq l\frac{\pi}{n} + \frac{\exp(-c^{n+1})}{n},$$

$$l\frac{\pi}{n} + \pi - \frac{\exp(-c^{n+1})}{n} \leqq \arg x_j \leqq l\frac{\pi}{n} + \pi + \frac{\exp(-c^{n+1})}{n},$$

and

$$x_j \in c_j,\ j = 1, 2,\ l = 0, \ldots, n-1, \tag{3.10 e}$$

where

$$\left.\begin{aligned} a_1 &= c \\ a_2 &= \ln c. \end{aligned}\right\} \tag{3.10 f}$$

and

Combining equations (3.8), (3.9) and (3.10) leads to

$$\oint_{x_1 \in c_1} dx_1 \oint_{x_2 \in c_2} dx_2 \, (c \ln c)^n (g(x_1, c) g(x_2, c))^{1-1/n} (\omega(c, g(x_1, c)) \omega(\ln c, g(x_2, c)))^{-1}$$

$$\times h_1(x_1, x_2, \tau_1, \tau_2, \eta, c) = \int_{x_1 \in s_1} dx_1 \int_{x_2 \in s_2 + s_4} dx_2 (g(x_1, c) g(x_2, c))^{-1/n}$$

$$\times \{-K(x_1, x_2, \tau_1, 0, \tau_1, \tau_2, c)$$

$$\times (1 + \varepsilon_1(x_1, x_2, c)) + K(x_1, x_2, \tau_1, \eta, \tau_1, \tau_2, c)(1 + \varepsilon_2(x_1, x_2, c))$$

$$+ K(x_1, x_2, 0, 0, \tau_1, \tau_2, c)(1 + \varepsilon_3(x_1, x_2, c))$$

$$- K(x_1, x_2, 0, \eta, \tau_1, \tau_2, c)(1 + \varepsilon_4(x_1, x_2, c))\}(\omega(\ln c, g(x_2, c)))^{-1}$$

$$+ \int_{x_1 \in s_1} dx_1 \int_{x_2 \in s_2 + s_4} dx_2 \{g(x_1, c) g(x_2, c)\}^{-1/n} \varepsilon_5(x_1, x_2, c)$$

$$+ \int_{x_1 \in s_3} dx_1 \int_{x_2 \in s_2 + s_4} dx_2 \, \varepsilon_6(x_1, x_2, \tau_1, c) \quad \int_{x_1, x_2} dx_1 \int_{\in L_1 L_2} dx_2 \varepsilon_7(x_1, x_2, c). \tag{3.11}$$

Repeated application of the procedure used in the proof of lemma 2 shows that

$$\varepsilon_p(x_1, x_2, c) \sim \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} a_{l,m}^{(p)}(c)(cx_1)^{-l}(\ln cx_2)^{-m}, \quad p = 1, \ldots, 4. \tag{3.12 a}$$

Moreover,

$$\varepsilon_5(x_1, x_2, \tau_1, c) = 0\{\exp(-c^{n-1/2})\}, \tag{3.12 b}$$

$$\varepsilon_6(x_1, x_2, \tau_1, c) = 0\{\exp(-c^{n-1/2}(1-\tau_1))\}, \tag{3.12 c}$$

$$\int_{x_1, x2} dx_1 \int_{\in L_1, L_2} dx_2 \varepsilon_7(x_1, x_2, c) = 0(1), \tag{3.12 d}$$

$$s_j = \left\{ x_j : x_j \in c_j, \, 2l \frac{\pi}{n} + \exp(-c^{n+1}) \leqq \arg x_j \leqq (2l+1) \frac{\pi}{n} - \exp(-c^{n+1}) \right\}, \tag{3.13 a}$$

$$s_{j+2} = \left\{ x_j : x_j \in c_j, (2l+1) \frac{\pi}{n} + \exp(-c^{n+1}) \leqq \arg x_j \leqq (2l+2) \frac{\pi}{n} - \exp(-c^{n+1}) \right\}, \tag{3.13 b}$$

if

$$j = 1, 2, \quad \text{and} \quad l = 0, 1, \ldots, n-1,$$

$$L_1 = \{x_1 \in c_1; x_2 \in c_2 - s_2 - s_4\} \tag{3.13 c}$$

and

$$L_2 = \{x_1 \in c_1 - s_1 - s_3; x_2 \in c_2\}. \tag{3.13 d}$$

The second integral on the right-hand side of equation (3.11) estimates the error made by replacing $(\omega(c, g(x_1, c)))^{-1}$ by unity if $x_1 \in s_1$, using equation (3.10); the third integral estimates the contribution to the integral of the parts $x_1 \in s_3$, $x_2 \in s_2 + s_4$; and the last integral estimates the contribution of the intervals of width $\exp(-c^{n+1})$ around the 'critical' points $x_1=\exp(il\pi/n)$ and $x_2=(il'\pi/n)$, $l=0,1,2,\ldots,2n-1$ and $l'=0,1,2,\ldots,n-1$.

We will now calculate *asymptotically* the contribution of the function h_2 to the integral (3.4). This becomes feasible using an ingenious device first introduced into complex analysis by Cauchy [14] (see also Picard [15] in his theory of Fourier series). First we observe that by Cauchy's theorem the integral

$$\oint_{x_1\in c_1} dx_1 \oint_{x_2\in c_2} dx_2\,(c\ln c)^n (g(x_1,c)g(x_2,c))^{1-1/n} \times(\omega(\ln c, g(x_2,c)))^{-1} h_2(x_1,x_2,\tau_1,\tau_2,\xi,\eta,c)=0 \qquad (3.14)$$

because, recalling lemma 1, the integrand of equation (3.14) is *entire within the domain* $|x_1|<1$ and continuous at and near the boundary. Subtracting equation (3.14) (i.e. zero) obviously does not change the value of the integral

$$I=\oint_{x_1\in c_1} dx_1 \oint_{x_2\in c_2} dx_2\,(c\ln c)^n (g(x_1,c)g(x_2,c))^{1-1/n} \times(\omega(c,g(x_1,c))\omega(\ln c, g(x_2,c)))^{-1} h_2(x_1,x_2,\tau_1,\tau_2,\xi,\eta,c), \qquad (3.15)$$

changing it into

$$I=-\oint_{x_1\in c_1} dx_1 \oint_{x_2\in c_2} dx_2\,(c\ln c)^n (g(x_1,c)g(x_2,c))^{1-1/n} \frac{\exp icg(x_1,c)}{(\exp icg(x_1,c)+1)} \times(\omega(\ln c, g(x_2,c)h_2(x_1,x_2,\tau_1,\tau_2,\xi,\eta,c). \qquad (3.16)$$

The reason for doing this is that the sampling function $(\omega(c,g(x_1,c)))^{-1} + \exp(icg(x_1,c))$ shows a complementary asymptotic behaviour with respect to $(\omega(c,x_1))^{-1}$, i.e. this new sampling function tends to zero in those regions where the former function tended to unity, and tends to minus unity where the former function tended to zero, as is obvious by subtracting unity from both sides of equations (3.10 a) to (3.10 *e*). This complementary behaviour is needed because, as is immediately clear from the asymptotic expansion (3.9), the function h_2 shows such a complementary behaviour with respect to h_1 on the unit circle $|x|=1$. Recalling that

$$-\exp(icg(x_1,c))(\omega(c,g(x_1,c)))^{-1}\equiv\omega(c,g(x_1,c))^{-1}-1, \qquad (3.17)$$

we obtain from equations (3.9), (3.10 *a*)–(3.10 *e*) and (3.15)–(3.17),

$$\begin{aligned}I=&\int_{x_1\in s_3} dx_1 \int_{x_2\in s_2+s_4} dx_2\,(g(x_1,c)g(x_2,c))^{-1/n}\{K(x_1,x_2,\tau_1,0,\tau_1,\tau_2,c)(1+\varepsilon_1(x_1,x_2,c))\\ &-K(x_1,x_2,\tau_1,\eta,\tau_1,\tau_2,c)(1+\varepsilon_2(x_1,x_2,c))-K(x_1,x_2,\xi,0,\tau_1,\tau_2,c)\\ &\times(1+\varepsilon_8(x_1,x_2,c))\\ &+K(x_1,x_2,\xi,\eta,\tau_1,\tau_2,c)(1+\varepsilon_9(x_1,x_2,c))\}\omega(\ln c,x_2)^{-1}\\ &+\int_{x_1\in s_3} dx_1 \int_{x_2\in s_2+s_4} dx_2\,(g(x_1,c)g(x_2,c))^{-1/n}\\ &\times\varepsilon_{10}(x_1,x_2,c)+\int_{x_1\in s_1} dx_1\int_{x_2\in s_2+s_4} dx_2\,\varepsilon_{11}(x_1,x_2,c)\\ &+\int_{x_1,x_2} dx_1\int_{\in L_1,L_2} dx_2\varepsilon_{12}(x_1,x_2,c). \end{aligned} \qquad (3.18)$$

The functions ε_p admit the asymptotic expansion

$$\varepsilon_p(x_1, x_2, c) \sim \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} a_{l,m}^{(p)} (cx_1)^{-l} (\ln cx_2)^{-m}, \quad p=8,9,$$

and

$$\varepsilon_{10}(x_1, x_2, c) = 0\{\exp(-c^{n1/2}(1-\xi+\tau_1))\},$$

$$\varepsilon_{11}(x_1, x_2, c) = 0(1)$$

$$\int_{x_1, x_2} dx_1 \int_{\in L_1, L_2} dx_2 \, \varepsilon_{12}(x_1, x_2, c) = 0\{\exp(-c^{n+1} + c^n)\}. \tag{3.19}$$

From the definitions (3.13 *a*)–(3.13 *d*) of the sectors s_1–s_4, we derive

$$\int_{x_1 \in s_1 + s_3} dx_1 \int_{x_2 \in s_2 + s_4} dx_2 - \oint_{x_1 \in c_1} dx_1 \int_{x_2 \in s_2 + s_4} dx_2 \, [(g(x_1, c) g(x_2, c))^{1/n}$$

$$\times \{K(x_1, x_2, \tau_1, 0, \tau_1, \tau_2, c)(1+\varepsilon_1(x_1, x_2, c)) - K(x_1, x_2, \tau_1, \eta, \tau_1, \tau_2)$$

$$(1+\varepsilon_2(x_1, x_2, c))\}(\omega(\ln c, g(x_2, c)))^{-1}] = 0\{\exp(-c^{n+1} + c^{n})\}. \tag{3.20}$$

From the power series (2.12) we derive

$$cg^{1/n}\left(\frac{z}{c}, c\right) = z + 0\left\{\frac{1}{c}\right\}, \quad \text{if } |z| \leqq 1, \tag{3.21}$$

and therefore

$$(\omega(\ln(c), x_2))^{-1} K(x_1, x_2, \tau_1, y_j, \tau_1, \tau_2, c)(1+\varepsilon_j(x_1, x_2, c)) = 0(1)$$

if

$$|x_1| = \frac{1}{c}, \quad x_2 \in s_2 + s_4, \quad \text{and} \quad y_1 = 0, \quad y_2 = \eta, \quad j=1,2. \tag{3.22}$$

Hence, by changing the contour $x_1 = c_1$ to $|x_1| = c^{-1}$ which, according to Cauchy's theorem can be done without changing the value of the second integral on the right-hand side of equation (3.20), combining equations (3.20) and (3.22) yields

$$\int_{x_1 \in s_1 + s_3} dx_1 \int_{x_2 \in s_2 + s_4} dx_2 - \oint_{|x_1| = c^{-1}} dx_1 \int_{x_2 \in s_2 + s_4} dx_1 [\{K(x_1, x_2, \tau_1, 0, \tau_1, \tau_2, c)$$

$$\times (1+\varepsilon_1(x_1, x_2, c)) - K(x_1, x_2, \tau_1, \eta, \tau_1, \tau_2, c)$$

$$\times (1+\varepsilon_2(x_1, x_2, c))\}(\omega(\ln c, x_2))^{-1}] = 0(1). \tag{3.23}$$

From equations (3.5) and (3.13 *a*)–(3.13 *d*), we obtain

$$(\omega(\ln c, g(x_2, c))^{-1} K(x_1, x_2, 0, y_j, \tau_1, \tau_2, c) = 0\{\exp(-c^{n-1/2})\}$$

if

$$x_1 \in s_1, \quad x_2 \in s_2 + s_4, \quad y_1 = 0, \quad y_2 = \eta, \quad \text{and} \quad j = 1, 2, \tag{3.24}$$

and

$$(\omega(\ln c, g(x_2, c)))^{-1}K(x_1, x_2, \xi, y_j, \tau_1, \tau_2, c) = 0\{\exp(-c^{-1/2}(\xi - \tau_1))\},$$

if

$$x_1 \in s_3, \quad x_2 \in s_2 + s_4, \quad y_1 = 0, \quad y_2 = \eta, \quad \text{and} \quad j = 1, 2. \tag{3.25}$$

Then, combining equations (3.4), (3.6), (3.7), (3.11), (3.12 *a*)–(3.12 *d*), (3.15), (3.18), (3.19), (3.23), (3.23), (3.24) and (3.25) yields

$$I(\zeta, \eta, \tau_1, \tau_2, c) = 0(1), \quad \text{if} \quad c \to \infty \quad \text{and} \quad \tau_1 < \xi, \tau_2 > \eta, \tag{3.26}$$

and hence

$$\lim_{c \to \infty} I(\xi, \eta, \tau_1, \tau_2, c) = 0, \quad \tau_1 < \xi, \tau_2 \eta. \tag{3.27}$$

We will now briefly indicate how the integral (3.4) can be evaluated if $\tau_1 < \xi, \tau_2 < \eta$. We therefore introduce four functions h_j, $j = 1, 2, 3, 4$:

$$h_1 = \int_0^{\tau_1} dy_1 \int_0^{\tau_2} dy_2\, K, \tag{3.28 a}$$

$$h_2 = \int_0^{\tau_1} dy_1 \int_{\tau_2}^{\eta} dy_2\, K, \tag{3.28 b}$$

$$h_3 = \int_{\tau_1}^{\xi} dy_1 \int_0^{\tau_2} dy_2\, K \tag{3.28 c}$$

and

$$h_4 = \int_{\tau_1}^{\xi} dy_1 \int_{\tau_2}^{\eta} dy_2\, K, \tag{3.28 d}$$

and an analysis completely analogous to the one developed above leads to

$$I(\xi, \eta, \tau_1, \tau_2, c) = \frac{1}{4\pi^2} \oint_{|x_1|=1} dx_1 \int_{|x_2|=1} dx_2\, (g(x_1, c) g(x_2, c))^{-1/n} + \mathrm{o}(1)$$

if

$$c \to \infty, \quad \tau_1 < \xi, \tau_2 < \eta, \tag{3.29}$$

where we used

$$K(x_1, x_2, y_1, y_2, \tau_1, \tau_2, c)\big|_{y_1 = \tau_1,\, y_2 = \tau_2} = 1 \tag{3.30}$$

and changed the contours c_1 and c_2 to $|x_1| = 1$ and $|x_2| = 1$, which can be done without changing the value of the integral at the right-hand side of equation (3.29) because the only singularities inside the contours c_1 and c_2 are the points $x_1 = 0$ and $x_2 = 0$. Then, because $g^{1/n}(z, c) = z + 0\{c^{-1}\}$ if $|z| = 1$ (see lemma 1), equation (3.29) leads to

$$\lim_{c \to \infty} I(\xi, \eta, \tau_1, \tau_2, c) = -\frac{1}{4\pi^2} \oint_{|x_1|=1} \frac{dx_1}{x_1} \oint_{|x_2|=1} \frac{dx_2}{x_2} + \mathrm{o}(1) \tag{3.31 a}$$

$$= 1, \ \tau_1 < \xi, \tau_2 < \eta. \tag{3.31 b}$$

A similar analysis, if $\tau_1 > \xi, \tau_2 < \eta, \tau_1 > \xi, \tau_2 > \eta$ leads to

$$\lim_{c \to \infty} I(\xi, \eta, \tau_1, \tau_2, c) = U(\xi - \tau_1) U(\eta - \tau_2). \tag{3.32}$$

We now come to the main theorem of this theory of generalized Fourier optics.

Theorem 2†

Let $\psi(y_1, y_2)$ be a complex function of bounded variation of the real variables y_1 and y_2 and defined in an *arbitrary* bounded domain D of the y_1, y_2 plane. Then, if for all complex values of x_1 and x_2

$$h(x_1, x_2) = \int_D \int dy_1 dy_2 \exp(iS(x_1, x_2, y_1, y_2))\psi(y_1, y_2), \tag{3.33}$$

equation (3.33) considered as an integral equation of the first kind with known function $h(x_1, x_2)$ can be solved uniquely for the unknown function $\psi(y_1, y_2)$ and

$$\psi(\tau_1, \tau_2) = \lim_{c \to \infty} \sum_j \sum_{j'} (g(x_{1j}, c)g(x_{2j'}, c))^{1-1/n}(g'(x_{1j}, c)g'(x_{2j'}, c))^{-1}$$
$$\times \exp(-iS(cx_{1j}, \ln cx_{2j'}, \tau_1, \tau_2))h(cx_{1j'} \ln cx_{2j'}), \tag{3.34}$$

where the function $g(z, c)$ has been defined in lemma 1 and the numbers x_{1j} and $x_{2j'}$, are the roots of the transcendental equations $\exp(ic^n x_{1j}^n) + 1 = 0$ and $\exp(i \ln^n cx_{2j'}^n) + 1 = 0$, whose moduli do not exceed unity.

Proof

The function $\psi(y_1, y_2)$ is of bounded variation, and therefore the two-dimensional Stieltjes integral

$$J(x_1, x_2) = \int_D \int d^2{}_{y_1, y_2}\psi(y_1, y_2) \tag{3.35}$$

exists.

Let the domain D be embedded within a rectangular domain D [27]:

$$D' = \{y_1, y_2 : \alpha \leqq y_1 \leqq \beta, \gamma \leqq y_2 \leqq \delta\}; \tag{3.36}$$

integrating equation (3.33) by parts then leads to

$$\begin{aligned} h(x_1, x_2) = {}& \psi(y_1, y_2) \int_\alpha^{y_1} dt_1 \int_\gamma^{y_2} dt_2 \exp(iS(x_1, x_2, t_1, t_2)) \Bigg|_{y_1=\alpha}^{y_1=\beta} \Bigg|_{y_2=\gamma}^{y_2=\delta} \\ & - \int_\alpha^\beta \left[\int_\alpha^{y_1} dt_1 \int_\gamma^{y_2} dt_2 \exp(iS(x_1, x_2, t_1, t_2))\, d_{y_1}(\psi(y_1, y_2)) \right]_{y_2=\gamma}^{y_2=\delta} \\ & - \int_\gamma^\delta \left[\int_\alpha^{y_1} dt_1 \int_\gamma^{y_2} dt_2 \exp(iS(x_1, x_2, t_1, t_2))\, d_{y_2}(\psi(y_1, y_2)) \right]_{y_1=\alpha}^{y_1=\beta} \\ & + \int_D \int d^2_{y_1, y_2}\psi(y_1, y_2) \int_\alpha^{y_1} dt_1 \int_\gamma^{y_2} dt_2 \exp(iS(x_1, x_2, t_1, t_2)). \end{aligned} \tag{3.37}$$

The Tannery series on the left-hand side of equation (3.3) converges boundedly for all values of y_1 and $y_2 \in D'$, and hence we can interchange, according to Lebesgues's

† The validity of equation (3.34) would be apparent from equation (3.3) if we were allowed to change limits and integrations. Using the relation $d/dx U(x) = \delta(x)$, the right-hand side of equation (3.34) would lead to

$$\int_D \int \frac{d}{dy_1}(U(y_1 - \tau_1)) \frac{d}{dy_2}(U(y_2 - \tau_2))\psi(y_1, y_2)\, dy_1 dy_2,$$

which equals $\psi(\tau_1, \tau_2)$. This theorem suggests that inversion might be possible for any eikonal.

dominated convergence theorem, limit and integration. Then, combining equations (3.3) and (3.37) leads to

$$\lim_{c\to\infty}\sum_j\sum_{j'}(g(x_{1j},c)g(x_{2j},c))^{1-1/n}(g'(x_{1j},c)g'(x_{2j},c))^{-1}$$

$$\times\exp(-iS(cx_{1j},\ln cx_{2j'},\tau_1,\tau_2))h(cx_{1j},\ln cx_{2j'})$$

$$=\psi(y_1,y_2)U(y_1-\tau_1)U(y_2-\tau_2)\Big|_{y_1=a}^{y_1=b}\Big|_{y_2=c}^{y_2=d}$$

$$-\int_a^b U(y_1-\tau_1)U(y_2-\tau_2)\,d_{y_1}\psi(y_1,y_2)\Big|_{y_2=c}^{y_2=d}$$

$$-\int_c^d U(y_1-\tau_1)U(y_2-\tau_2)\,d_{y_2}\psi(y_1,y_2)\Big|_{y_1=a}^{y_1=b}$$

$$+\int_a^b dy_1\int_c^d dy_2 d^2_{y_1,y_2}(\psi(y_1,y_2))U(y_1-\tau_1)U(y_2-\tau_2)=\psi(\tau_1,\tau_2). \quad (3.38)$$

Equation (3.34) is easily recognized as a generalization of the ordinary Fourier series. The sampling points x_{1j} and x_{2j} are the numbers

$$\left.\begin{aligned}
&x_{1j}=((2j+1)\pi)^{1/n}c^{-1}\exp\left(2\pi i\frac{n'}{n}\right),\quad j=0,\pm1,\pm2,\dots\\
&\text{and}\\
&x_{2j'}=((2j'+1)\pi^{1/n})(\ln c^{-1})\exp\left(2\pi i\frac{n'}{n}\right),\quad j'=0,\pm1,\pm2,\dots\\
&\qquad\qquad n'=0,1,2,\dots,n,
\end{aligned}\right\}\quad (3.39)$$

whose modulus does not exceed unity. If no 'mixed' terms like $x_1^{n-3}x_2\tau_1$ occur in S, i.e. if

$$S=x_1^n\tau_1+x_2^n\tau_2, \quad (3.40)$$

equation (3.34) leads to

$$\psi(\tau_1,\tau_2)=\lim_{c\to\infty}\sum_j\sum_{j'}(g(x_{1j},c)g(x_{2j'},c))^{1-1/n}(g'(x_{1j},c)\,g'(x_{2j'},c))^{-1}$$

$$\times\exp(-i(2j+1)\pi\tau_1-i(2j'+1)\pi\tau_2)\iint_D\exp(i(2j+1)\pi y_1$$

$$+i(2j'+1)\pi y_2)\psi(y_1,y_2)\,dy_1dy_2. \quad (3.41)$$

The summation has to be taken over all the values of j and j' specified by equations (3.39). Equations (2.14), (2.15), (2.16) and (2.17) show that

$$\lim_{c\to\infty}(g(z,c))^{1-1/n}=z^{n-1}\quad\text{and}\quad\lim_{c\to\infty}g'(z,c)=nz^{n-1}. \quad (3.42)$$

Combining equations (3.41) and (3.42) shows that equation (3.41) leads to an ordinary Fourier series if we interchange limit and summation, which is allowed by a theorem due to Bromwich [28]:

$$\psi^{(1)}(\tau_1,\tau_2)=\sum_j\sum_{j'}\exp(-i2j\pi\tau_1-i2j'\pi\tau_2)$$

$$\times\iint_D \exp(i2j\pi y_1+i2j'\pi y_2)\psi^{(1)}(y_1,y_2)\,dy_1\,dy_2, \quad (3.43)$$

if

$$\psi^{(1)}(\tau_1,\tau_2)=\psi(\tau_1,\tau_2)\exp(i\pi\tau_1+i\pi\tau_2). \quad (3.44)$$

The situation is different if mixed terms occur in the eikonal S. For instance, it is not possible to change limit and summation because the general term of the series thus obtained would tend to infinity. The series would therefore diverge. It is explained in [28] why the original series converges, while the general term does not tend to zero: each term of the series depends on the upper limit of the summation. However, the structure of the series (3.34) is very similar to the Fourier series (3.43). The general term of (3.34) is obtained by multiplying the function $h(x_1,x_2)$ by $\exp(-iS(x_1,x_2,\tau_1,\tau_2))$, taking x_1 and x_2 equal to one of the points cx_{1j} and In $(c)x_{2j'}$ respectively and multiplying this result by the constant $(g(x_{1j},c)g(x_{2j'},c))^{1-1/n}(g'(x_{1j},c)g'(x_{2j'},c))^{-1}$.It is this last multiplication which is the main difference between the Fourier series (3.43) and the series (3.34).

3. Discussion

In this paper we have considered a reconstruction problem concerning the determination of a wavefunction ψ at a plane $z=z_0$ from the values of a wavefunction h (which is uniquely determined by the wavefunction ψ), at a plane $z=z_1$. We assumed that the optical instrument in which the image formation takes place suffers from an arbitrary number of aberrations and showed that, generalizing the theory of Fourier optics, the wavefunction ψ is uniquely determined by h.

Our theory of generalized Fourier optics applies to wavefunctions rather than to intensity distributions which are the only measurable quantities in electron and light optics. However, if the theory of Fourier optics can be applied, several methods by which the phase of the wavefunction can be determined from its modulus (intensity) have been proposed and tested [16]. (For an extensive survey of the literature up to 1972 see [17–25].)

Our theory could therefore be applied if, in a region near the image plane, the theory of Fourier optics is valid and an arbitrary number of aberrations have to be taken into account outside this region.

Although we have developed a Fourier series type of inversion theorem (equation (3.34)), this series might, due to the occurrence of complex-valued sampling points, exhibit a very unstable behaviour. Therefore, for numerical purposes, we probably have to rely on other algorithms, but the theory of generalized Fourier optics at least proves the uniqueness of the solution.

The importance of the theory of generalized Fourier optics which, in view of the present, highly corrected, light optical instruments seems to be of rather academic interest, is at least twofold:

(1) The theory provides us with the natural set of functions with which, as when the set of complex exponentials $\{\exp(inx)]$ was used for band-limited

functions, the calculations concerning the information content of an image blurred only by spherical aberration and defocusing [3] can be extended to images blurred by an arbitrary number of aberrations.

(2) The theory of generalized Fourier optics shows that it is not necessary to dispose of the information contained in the electrons intercepted by the small-aperture diaphragms currently used in electron microscopes. The main reason for using such diaphragms is to reduce the influence of aberrations of bad electromagnetic lenses on the imaging properties of the microscope, making the application of the techniques of Fourier optics feasible. However, because generalized Fourier optics gives a procedure for reversing the order of image formation in a microscope suffering from an arbitrary number of aberrations, it is not necessary to use small apertures and we can make use of the information contained in electrons which at present do not contribute to the image.

Acknowledgment

The author wishes to thank Dr. H. A. Ferwerda for several critical comments.

References

[1] DUFFIEUX, P. M., 1946, *Les Integrals de Fourier et ses Applications a l'Optique* (Rennes).
[2] BORN, M., and WOLF, E., 1965, *Principles of Optics* (New York: Pergamon).
[3] FELGETT, P. B., and LINFOOT, E. H., 1955, *Phil. Trans. R. Soc.* A, **247,** 369.
[4] LANDE, A., 1928, *Handbuch der Physik* (Berlin: Springer), Chap. 9.
[5] WOLTER, H., 1958, *Physica, 's Grav.*, **24,** 457; 1959, *Arch. elekt. Übertr.*, **13,** 393; 1961, *Progress in Optics*, Vol. 1 (Amsterdam: North Holland).
[6] TORALDO DI FRANCIA, G., 1969, *J. opt. Soc. Am.*, **59,** 799.
[7] HOENDERS, B. J., 1973, *Optik*, **37,** 542.
[8] LALOR, E., 1968, *J. Math. Phys.*, **9,** 2001.
[9] SHEWELL, J. R., and WOLF, E., 1968, *J. opt. Soc. Am.*, **58,** 1596.
[10] FERWERDA, H. A., and HOENDERS, B. J., 1974, *Optik*, **39,** 317.
[11] HILBERT, D., 1912, *Grundzüge einer Allgemeinen Theorie der Linearen Integralgleichungen* (Leipzig, Berlin: Teubner).
[12] WHITTACKER, E. T., and WATSON, G. N., 1965, *A Course of Modern Analysis* (Cambridge University Press).
[13] WATSON, G. N., 1914, *Complex Integration and Cauchy's Theorem*, Cambridge Mathematical Tracts No. 15. For a generalization of Watson's results, see POLLARD, S., 1923 *Proc. Lond. math. Soc.*, **21,** 456.
[14] CAUCHY, A. L., 1827, *Exercises de Mathématique* Vol. II, p. 603 (see also *Oeuvres* (2), **7,** 393). For a rigorous mathematical analysis see HARNACK, A., 1888, *Math. Annln*, **32,** p. 175.
[15] PICARD, E., 1905, *Traité d'Analyse*, Vol. 2, second edition (Paris: Gauthiers Villard), p. 179.
[16] WOLF, E., 1962, *Proc. phys. Soc.*, **80,** 1296.
[17] MANDEL, L., and KOHLER, D., 1973, *J. opt. Soc. Am.*, **63,** 126.
[18] NUSSENZVEIG, H. M., 1967, *J. Math. Phys.*, **8,** 561.
[19] ROMAN, P., and MARATHAY, A. S., 1963, *Nuovo Cim.*, **30,** 1452.
[20] FERWERDA, H. A., 1976, *Optics Commun.*, **19,** 54.
[21] HUISER, A. M. J., VAN TOORN, P., and FERWERDA, H. A., 1977, *Optik*, **47,** 1.
[22] VAN TOORN, P., and FERWERDA, H. A., 1976, *Optica Acta*, **23,** 469; 1977, *Optik*, **47,** 123.
[23] HOENDERS, B. J., 1975, *J. Math. Phys.*, **16,** 1719.
[24] MISELL, D. L., 1974, *J. Phys.* D, **7,** 69, 832. See also MISELL, D. L., BURGE, R. E., and GREENAWAY, A. H., 1974, *J. Phys.* D, **7,** 27.
[25] GERCHBERG, R. W., and SAXTON, W. O., 1971, *Optik*, **34,** 275.

[26] The conditions that $a_{n,0,1,0}$ and $a_{0,n,0,1}$ are zero does not affect the generality of equation (2.22). If these numbers are not zero they can be made equal to zero by a suitable linear transformation in the y_1–y_2 plane. After $a_{n,0,1,0}$ and $a_{0,n,0,1}$ have been made equal to zero, we choose the rectangular domain $\alpha \leqslant y_1 \leqslant \beta, \gamma \leqslant y_2 \leqslant \delta$ as the domain of integration. (See also the proof of theorem 3, equation (3.36).)

[27] It is always possible to find a suitable scale transformation of x_1, x_2, y_1 and y_2 such that $|\alpha|$, $|\beta|$, $|\gamma|$ and $|\delta|$ are smaller than 1 and that a_1 and a_2 are equal to unity.

[28] BROMWICH, T. J. J. A., 1965, *An Introduction to The Theory of Infinite Series* (London: Macmillan), § 174. The theorem of § 174 is also valid if double series are considered, and shows, together with equation (2.8), that it is allowed to interchange summation and limit.